Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Molecular Immunology 40 (2003) 247–260
Molecular Immunology
www.elsevier.com/locate/molimm

Structure, genomic organization, and phylogenetic implications of six new VH families in the channel catfish

Feixue Yang [a], Tereza Ventura-Holman [a], Geoffrey C. Waldbieser [b], Craig J. Lobb [a,*]

[a] *Department of Microbiology, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216-4505, USA*
[b] *US Department of Agriculture, Catfish Genetics Research Unit, Thad Cochran National Warmwater Aquaculture Center, Stoneville, MS 38776, USA*

Received 24 March 2003; received in revised form 26 May 2003; accepted 27 May 2003

Abstract

To define members of previously unknown VH gene families, a channel catfish immunoglobulin heavy chain cDNA library was constructed and screened with probes specific for the seven known catfish VH families. Reiterative screening and sequence studies defined six new VH families, designated VH8–VH13, which brings the total number of VH families in the catfish to 13. This is the highest number of VH families presently defined in a lower vertebrate. Sequence comparisons indicate there is extensive diversity between members of different families with the greatest variability encoded within the complementarity determining regions. Genomic libraries were screened, and germline VH segments representing each of these new families were identified. The VH segments are closely linked and interspersed with members of different VH families. Each of these germline gene segments shared characteristic structural features: an upstream region that contained transcriptional regulatory elements, a leader sequence split by a short intron, an open reading frame encoding readily identified framework and complementarity determining regions, and a terminal recombination signal sequence consisting of a consensus heptamer, a 22–24 bp spacer with conserved 5′- and 3′-ends, and a consensus A-rich nonamer. Southern blot analyses estimate the number of members within these new families ranges from small (2–7 members in VH9, VH10, and VH12) to medium (9–13 members in VH8, VH11, and VH13). Thus, there are between 165 and 200 germline VH segments represented by these combined 13 families with present analyses indicating that perhaps one-half of these are pseudogenes. Phylogenetic comparisons indicate that members of these different catfish VH families cluster within Groups C and D of vertebrate VH genes. These analyses also indicate that Group D is represented by two different branches and both branches include VH families from different lineages of bony fish that diverged early in vertebrate phylogeny.

Keywords: Evolution; Gene rearrangement; Immunoglobulin heavy chain variable region; Multigene family

1. Introduction

The genomic complexity of three different sets of variable region gene segments provides the inherent structural diversity of immunoglobulin (Ig) heavy (H) chains. These gene segments, referred to as VH, DH, and JH, undergo somatic recombination to form the rearranged V–D–J. This process is mediated by recombinase enzymes which recognize specific recombination signal sequences (RSS) which flank each of these gene segments. Combinatorial diversity between these gene segments is further extended by the processes of junctional diversity and somatic mutation. The H chain V region is about 110 amino acids in length and is composed of four framework (FR) regions and three complementarity determining regions (CDR). The first three FR regions and the first two CDR regions are encoded by VH segments, the CDR3 is encoded by contributions from VH, DH, and JH segments, and FR4 is encoded by JH segments (reviewed by Max, 1999). Comparisons of the nucleic acid similarity of VH genes have defined gene families wherein members of the same VH family share greater than 80% nucleic acid similarity. The similarity between members of different families is generally less than 70% (Brodeur et al., 1988; Kabat et al., 1991).

In the catfish, earlier studies had shown that there are >100 VH gene segments that are organized into seven different VH families. These gene segments are closely linked with an average distance between segments of about 3 kb and are located upstream of the DH and JH segments (Ghaffari and Lobb, 1991; Ventura-Holman et al., 1994; Ventura-Holman et al., 1996). The DH locus, identified through approaches that examined the excision products of DH–JH recombination events, comprises at least three DH segments and is

* Corresponding author. Tel.: +1-601-984-1700; fax: +1-601-984-1708.
E-mail address: clobb@microbio.umsmed.edu (C.J. Lobb).

0161-5890/$ – see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0161-5890(03)00143-3

located about 9 kb upstream of the nine segments that compose the JH locus (Hayman et al., 1993; Hayman and Lobb, 2000). The single-copy Cμ gene is located about 1.8 kb downstream from the JH segments and encodes the H chain C region of the predominant, if not the exclusive, serum Ig and Ab of the catfish (Lobb, 1985; Ghaffari and Lobb, 1989a,b; Ghaffari and Lobb, 1992). Recent studies have also shown that the catfish IgH locus underwent a massive internal duplication resulting in two linked gene clusters, with the duplicated upstream cluster containing VH segments representing multiple families, a JH segment, a germline-joined V–D–J segment, and pseudogene C region exons (Ghaffari and Lobb, 1999; Ventura-Holman and Lobb, 2002; Bengten et al., 2002).

During studies to characterize germline VH segments there were regions within genomic clones that hybridized under relaxed, but not stringent conditions, with known VH family specific probes. These genomic analyses suggested that uncharacterized VH families may exist. To address this hypothesis we quantitatively analyzed an Ig H chain cDNA library derived from an adult catfish to determine if additional VH families were expressed, and if so to characterize germline VH segments representing these new families. These analyses have led to this report which defines six additional VH families and provides important new insights into the explosion of VH structural diversity which occurred early in vertebrate phylogeny at the level of the bony fishes.

2. Materials and methods

2.1. Construction of a channel catfish immunoglobulin heavy chain cDNA library

Total RNA was extracted from the spleen of an individual adult outbred channel catfish (*Ictalurus punctatus*) using Trizol (Invitrogen, Carlsbad, CA), and 2 μg of RNA was used to synthesize doubled-stranded cDNA using a 5′-RACE kit (Invitrogen) according to the manufacturer's protocols. Briefly, first strand cDNA synthesis was initiated with a primer corresponding to the Cμ2 domain of the catfish H chain, and the product was tailed with poly-C at the 5′-end. Following RNAse digestion and spin column purification, the first strand product was amplified with a primer specific for the Cμ1 domain and the abridged adapter primer provided in the kit. PCR amplification was conducted using Taq-polymerase (Invitrogen) with the following amplification parameters: 1 min at 94 °C, 45 s at 52 °C, and 45 s at 72 °C for 30 cycles with the terminal cycle given a 7 min final extension time. The PCR products were analyzed in a agarose gel, and fragments estimated to be between 450 and 550 bp in length were excised. The amplicons were ligated into the T/A cloning vector pCR2.1, and Top10F′ competent cells (Invitrogen) were used for transformation. Following overnight incubation, individual colonies were transferred to master plates for use in subsequent analyses.

2.2. In situ colony hybridization and sequence analyses

Five hundred sixty colonies from the splenic Ig H chain cDNA library, as well as control colonies representing each of the seven known channel catfish VH families, were arrayed on LB agar plates, and following overnight culture, replicate lifts of the colonies were made using Nytran membranes (S&S, Keene, NH). Radiolabeled probes specific for each of the seven known channel catfish VH families were derived by PCR using the forward primers, reverse primers and plasmid templates shown in Table 1. The forward and reverse primers generally corresponded to areas within the encoding FR1 and FR3 regions, respectively, of the template. Hybridization analyses were performed under high stringency conditions (Ghaffari and Lobb, 1989b) and under these conditions the probes did not cross hybridize. Clones that likely represented new VH families were sequenced with ABI PRISM BigDye Terminators chemistry on an ABI PRISM 3700 DNA Analyzer (Applied Biosystems, Foster City, CA) in the USDA, ARS, Mid-South Area Genomic Laboratory. Sequences were derived using M13 forward and reverse primers, and all sequences were manually verified. The sequences of cDNA clones reported in this paper have

Table 1
PCR primers and templates used to derive channel catfish VH family specific probes

VH family	Forward primer	Reverse primer	Template (accession number)
VH1	5′-TGGGTGAAGCAGAACTCA-3′	5′-GCAGAGACTGGGCAAAG-3′	NG70 (M27230)
VH2	5′-GTTGCTTCCTATGTGCAT-3′	5′-GTCCTCGAATTGTTATTGT-3′	NG41 (M58671)
VH3	5′-CTCTTATCGAGTCTGATT-3′	5′-CTTCTTACTGTTGTCTCT-3′	NG21 (M58669)
VH4	5′-TCAGGAGATTCTTTAACC-3′	5′-TCTCAGAGTCACCGTACT-3′	NG10 (M58668)
VH5	5′-TGCTGCTGCTGGCTCTC-3′	5′-CTGCTGGACGTGTCTCGA-3′	NG66 (M58674)
VH6	5′-ATCCACACTCATAACCCCTG-3′	5′-CATGTTCTGCCCTGTTAGTG-3′	VH6.1 (U09724)
VH7	5′-TTTGCAGTGCTGAGATCAG-3′	5′-CTCTTGGCCTCTAAGTAC-3′	VH7E (AY238375)
VH8	5′-GTTGTGGATGTCAGAC-3′	5′-CTTCTTACTGTTGTCTCT-3′	VH8_A1 (AY238376)
VH9	5′-GGTTCAGCCTCCAGTCAT-3′	5′-CAGTTGTGTCCTCAGACT-3′	1G11AVH9 (AY238368)
VH10	5′-GTGTTGGCGGTGTGGAT-3′	5′-CAGCCGTGTCTCAGTCT-3′	1G07AVH10 (AY238370)
VH11	5′-TGTGCATGGCATCAGTCT-3′	5′-ACCAGGCTGCTAATATCT-3′	1F10AVH11 (AY238372)
VH12	5′-TCTCCTGCTGTTTCTGAC-3′	5′-CTGTGCTGCTGGAGACAT-3′	2G09AVH12 (AY238373)
VH13	5′-TATTCCCTCTGGTGCAGT-3′	5′-CACTCAAACCAGTCAGCT-3′	2G11AVH13 (AY238374)

been assigned the accession numbers AY238358 through AY238374.

2.3. Genomic analyses

Genomic clones containing members of previously unknown channel catfish VH gene families were derived by three approaches. VH8.1 was initially identified as a restriction fragment within lambda clone G2.3 which hybridized with a VH7 probe under relaxed stringency conditions. In addition a cDNA clone VH8_A1, derived from another individual channel catfish, was identified which contained an expressed member of VH8 (Ghaffari and Lobb, unpublished, accession number AY238376). Secondly, 36 plaque-purified genomic DASH II lambda clones, previously known to contain at least one member of the VH1–VH5 channel catfish families (Ventura-Holman et al., 1994) were screened by PCR approaches for members of the VH9–VH13 families using the primer pairs and templates shown in Table 1. By this approach, members of VH10, VH11, and VH13 were identified in phage clones vh2c, vh3f, and vh3a. Lastly, genomic DNA phage libraries, previously constructed in lambda DASH II (Ghaffari and Lobb, 1992), were screened under high stringency conditions with probes specific for VH9 and VH12, and selected positive clones were plaque-purified. Each of these identified clones was then restricted with different enzymes, and Southern blots were hybridized under high stringency conditions with the family specific probe used in clone identification (Ghaffari and Lobb, 1991). The positive restriction fragments were subcloned into pUC18 or pUC19 and sequenced with Sequenase (USB, Cleveland OH) using vector primers as well as various internal primers. The sequences of genomic VH members representing VH8–VH13 were designated as VH8.1, VH9.1, VH10.1, VH11.1, VH12.1, and VH13.1, respectively, and have been submitted to GenBank and assigned the accession numbers AY238377 through AY238382, respectively.

Genomic Southern blot analyses were conducted under high stringency conditions using the above probes with genomic DNA obtained from the nucleated erythrocytes of an individual outbred channel catfish using methods previously described (Ghaffari and Lobb, 1989b).

2.4. Phylogenetic analyses

The predicted amino acid sequences of VH segments from representative higher and lower vertebrates were aligned

Table 2
VH and outgroup sequences from different vertebrates used for phylogenetic analysis

Species	Gene family (sequence name, accession number)
Mammals	
Human (*Homo sapiens*)	HumVH1(VH1-2, AB019441), HumVH2(VH2-5, AB019440), HumVH3(VH3-9, AB019440), HumVH4(VH4-4, AB019441), HumVH5(VH5-51, AB019438), HumVH6(VH6-1, AB019441), HumVH7(VH7-81, AB019437)
Mouse (Mus musculus)	MusVH1(VHF102, AF305910), MusVH2(VOX-1, U53526), MusVH3(VH36-30, K01569), MusVH4(VHGa155.1, X01437), MusVH5(VH7183.23b, AF120463), MusVH6(VH22.1, X03398), MusVH7(V1, J00538), MusVH8(CB17H-1, U23019), MusVH9(VGK4, L14366), MusVH10(VH10.2a, AF064445), MusVH11(NW_000057, 379277-378969), MusVH12(NW_000057, 692192-691892), MusVH13(NW_000057, 683429-683130), MusVH14(vhsm7-13, X55934), MusVH15(NW_000057, 890737-890426)
Lower vertebrates	
Chicken (*Gallus domesticus*)	ChkVH1(VH1, M30338)
Caiman (*Caiman crocodylus*)	CaimanVH1 (C3, M12768)
African toad (*Xenopus laevis*)	XeVH1(XeH1, Y00380), XeVH2*(XeVH2, M24674), XeVH3*(XeVH3, M24675), XeVH4(Xe4, X56858), XeVH5(Xe5, X56859), XeVH6(Xe6, X56860), XeVH7(Xe7, X56861), XeVH9(Xe9, X56863), XeVH10(Xe10, X56864), XeVH11(Xe11, X56865)
Channel catfish (*Ictalurus punctatus*)	CcfVH1(VH1.1, U09719), CcfVH2*(NG41, M58671), CcfVH3(VH3.1, U09721), CcfVH4(VH4.1, U09722), Ccf VH5*(NG66, M58674), CcfVH6(VH6.1, U09724), CcfVH7*(1C09AVH7, AY238364), CcfVH8(VH8.1, AY238377), CcfVH9(VH9.1, AY238378), CcfVH10*(2G01AVH10, AY238369), CcfVH11*(1F10AVH11, AY238372), CcfVH12(VH12.1, AY238381), CcfVH13*(2G11AVH13, AY238374)
Rainbow trout (Oncorhynchus mykiss)	RbtVH1(RTVH431, M57442), RbtVH2*(X65262), RbtVH3*(X81510), RbtVH4*(L28744), RbtVH5*(X81513), RbtVH6*(X81497), RbtVH7*(L28747), RbtVH9*(X81507), RbtVH10*(X81508), RbtVH11*(X81511)
Little skate (*Raja erinacea*)	SkateVH (Re102, X16146)
Horned shark (*Heterodontus francisci*)	Horned shark VH (1320, X13448)
Outgroups	
Human (*Homo sapiens*)	HumVpre-B (VPB6, M34927)
Horned shark (*Heterodontus francisci*)	Horned shark VL (HFL122, X15316)

Germline VH sequences were used for phylogenetic analyses when available, otherwise cDNA sequences (marked by an asterisk) were used.

with the predicted amino acid sequences of catfish VH genes, and regions assigned to either FR or CDR regions according to Kabat et al. (1991). Because of sequence variability, length variation, and the large number of insertions and deletions required to align the CDR1 and CDR2 regions, the CDR regions were omitted. The sequence of the combined FR1–FR3 regions were then assembled, and the sequences aligned. The lack of sufficient VH germline sequences from lower vertebrates required that some cDNA sequences be used in phylogenetic analyses. Because earlier studies indicated that gene recombination between V–D–J segments in the channel catfish can extend into FR3 (Hayman and Lobb, 2000), the aligned FR1–FR3 sequences were terminated in FR3 at the end of the hallmark Y–Y–C found in virtually all VH sequences. The phylogenetic analysis of the amino acid sequences of the combined FR1–FR3 regions were conducted using the minimum evolution method with the Poisson correction distance (Rzhetsky and Nei, 1992; Nei and Kumar, 2000). The confidence probabilities on the interior branch lengths were calculated with a *t*-test with bootstrap resampling set to 1000 using the software suite MEGA2 (Kumar et al., 2001). The VH sequences representing the different VH family members used in these phylogenetic analyses are shown in Table 2. Representative sequences from rainbow trout VH8 and *Xenopus* VH8 were not included in these analyses because insertion/deletion events within FR regions made alignments with other vertebrate sequences tenuous.

3. Results

3.1. Analysis of a cDNA library derived from the spleen of an individual adult channel catfish

An Ig H chain cDNA library was constructed from RNA derived from the spleen of an individual adult channel catfish, and 560 randomly chosen clones from this library were sequentially screened by hybridization with specific probes for each of the seven known channel catfish VH gene families. By this approach, 424 clones could be identified as containing cDNA sequences which hybridized under high stringency conditions with these VH family specific probes. These results indicated that the majority of the cDNA sequences in this library (75.7%) represented expressed members of the known VH families and, importantly, suggested that additional VH families were expressed.

To identify sequences potentially representing expressed members of previously unknown VH families, a three-step approach was taken. First, the inserts from various non-hybridizing clones were sequenced. Secondly, probes were derived from these sequences that spanned the encoding FR1–FR3 regions, and dot-blot analyses were used to confirm that these probes did not cross-hybridize with plasmid inserts containing members of known VH families. Lastly, the clones were re-screened to identify which clones hybridized with these newly-derived probes, and clones of interest were sequenced. By this sequential approach the inserts contained within all 560 clones were classified, and clones containing six new groups of V region sequences were identified. Additional levels of complexity within the adult expressed repertoire will be reported subsequently (manuscript in preparation).

The coding region extending through the end of FR3 from clones representing these six groups of cDNA inserts were aligned with known sequences representing VH families 1–7. The nucleotide and predicted amino acid sequences are shown in Figs. 1 and 2. The V regions each exhibited the key structural coding features of known VH genes. These features included the conserved Leu at −10 or −11 in the leader sequence, the V region intradomain cysteines, the conserved tryptophan located at the beginning of FR2, and the conserved terminal FR3 coding sequence Y–Y–C–A–R. The nucleotide sequence similarities in the VH regions of these clones were compared, and the analyses confirmed that the new VH sequences could be placed into six different groups consistent with the designation of different VH families. In each of these comparisons, members within the same VH family shared greater than 80% nucleotide sequence similarity, and members of different VH families shared 70% or less nucleotide sequence similarity (Fig. 3). These families were designated VH8–VH13 with the prototype full-length cDNA sequence designated by the inserts in clones 3G01AVH8, 1G11AVH9, 1G07AVH10, 1F10AVH11, 2G09AVH12, and 2G11AVH13, respectively.

3.2. Coding region diversity between members of different catfish VH families

The inspection of the nucleic acid and amino acid alignments in Figs. 1 and 2 indicates that sequence diversity between members of different VH families is generally exhibited throughout the coding regions. To determine the extent of sequence diversity between members of these different families, the sequence similarities in the FR and CDR regions extending through the end of FR3 were compared (Table 3). These comparisons showed that FR2 regions were the most conserved with an average nucleotide similarity of 70%. The generally conserved sequences W–I–R–Q and E–W–I–G were, respectively, encoded at the beginning and the end of the FR2 regions. The FR3 regions were generally more conserved than the FR1 regions in the different families; the overall nucleotide similarity in FR3 regions ranged from 45 to 86%, whereas the nucleotide similarities in FR1 regions ranged from 38 to 68% (the average nucleotide similarities were 61% for FR3 and 51% for FR1). The nucleotide similarities in FR1 regions were generally dispersed, whereas the nucleotide similarities in FR3 regions were principally localized at the end of the regions and encoded the generally conserved 10 amino acid motif E–D–T–A–V–Y–Y–C–A–R. The lowest similarities defined in these comparisons were between

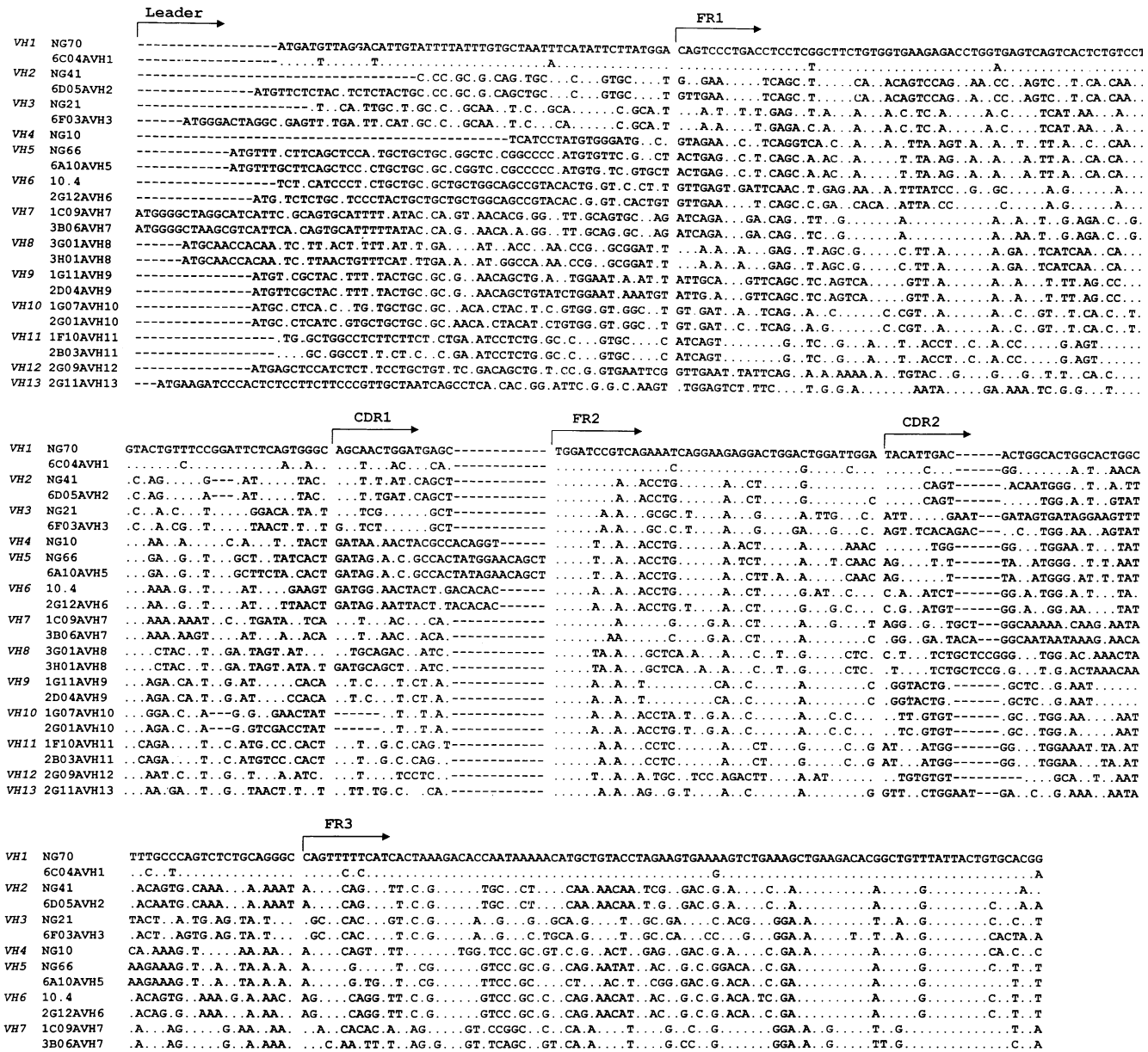

Fig. 1. Multiple alignment of the nucleotide sequences from cDNA clones representing the 13 different channel catfish VH families. The demarcations of the leader, framework (FR), and complementarity determining regions (CDR) are shown. The sequences are aligned with the NG70 VH1 sequence and nucleotide identities are indicated by dots whereas gaps introduced to maximize homology are indicated by dashes.

the CDRs. Nucleotide comparisons showed that CDR regions exhibited extensive variation between families; the similarities in the CDR1 regions ranged from 0 to 73%; whereas, the similarities in the CDR2 regions ranged from 13 to 67%. The number of codons in the CDRs also varied between these families. In CDR1 the number of codons ranged from 3 to 9, and in CDR2 the number of codons ranged from 15 to 18. In addition to CDR length variations, perhaps the most unusual feature was the presence of cysteine within both the CDR1 and CDR2 in members of the VH6, VH9, VH10, and VH12 families (see Section 4). Thus, these analyses indicate that there is extensive coding region diversity between members of these different catfish families.

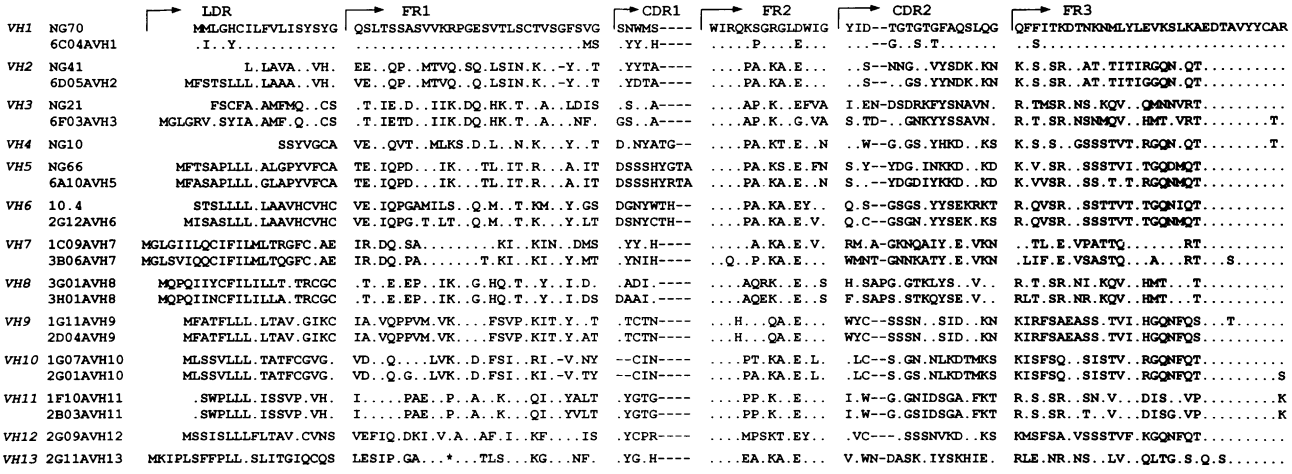

Fig. 2. Multiple alignment of the predicted amino acid sequences from cDNA clones representing 13 different channel catfish VH families. The sequences are aligned with the NG70 VH1 sequence and amino acid identities are indicated by dots whereas gaps introduced to maximize homology are indicated by dashes.

		VH1		VH2		VH3		VH4	VH5		VH6		VH7		VH8		VH9		VH10		VH11		VH12
		NG70	6C04 AVH1	NG41	6D05 AVH2	NG21	6F03 AVH3	NG10	NG66	6A10 AVH5	10.4	2G12 AVH6	1C09 AVH7	3B06 AVH7	3G01 AVH8	3H01 AVH8	1G11 AVH9	2D04 AVH9	1G07 AVH10	2G01 AVH10	1F10 AVH11	2B03 AVH11	2G09 AVH12
VH1	NG70																						
	6C04AVH1	90 (89)																					
VH2	NG41	55 (44)	56 (46)																				
	6D05AVH2	55 (46)	57 (48)	93 (91)																			
VH3	NG21	58 (40)	59 (42)	52 (37)	52 (36)																		
	6F03AVH3	57 (43)	57 (41)	52 (36)	54 (38)	85 (75)																	
VH4	NG10	52 (44)	53 (46)	62 (60)	66 (64)	50 (36)	49 (40)																
VH5	NG66	50 (38)	51 (39)	59 (51)	61 (51)	50 (40)	48 (41)	70 (60)															
	6A10AVH5	51 (40)	52 (42)	59 (55)	60 (56)	50 (40)	48 (40)	69 (61)	92 (88)														
VH6	10.4	49 (34)	52 (38)	62 (56)	64 (59)	49 (44)	47 (41)	65 (61)	67 (52)	64 (56)													
	2G12AVH6	53 (38)	53 (40)	64 (58)	66 (60)	52 (45)	52 (45)	67 (66)	69 (61)	66 (63)	88 (68)												
VH7	1C09AVH7	59 (52)	65 (59)	50 (41)	49 (41)	54 (41)	52 (38)	50 (36)	50 (38)	51 (40)	46 (41)	49 (41)											
	3B06AVH7	58 (46)	64 (54)	52 (41)	52 (41)	50 (35)	48 (33)	53 (39)	51 (41)	52 (41)	48 (40)	51 (42)	87 (75)										
VH8	3G01AVH8	56 (48)	60 (49)	49 (37)	50 (38)	67 (58)	68 (58)	51 (40)	50 (39)	51 (39)	49 (42)	49 (41)	51 (41)	91 (40)									
	3H01AVH8	57 (52)	61 (52)	50 (40)	50 (42)	68 (59)	68 (61)	51 (42)	49 (40)	50 (38)	50 (42)	50 (42)	52 (42)	49 (40)	93 (89)								
VH9	1G11AVH9	55 (38)	53 (36)	56 (50)	58 (51)	45 (30)	46 (28)	59 (44)	58 (43)	59 (45)	58 (44)	60 (48)	56 (39)	59 (44)	50 (34)	49 (33)							
	2D04AVH9	55 (38)	53 (35)	56 (50)	58 (51)	45 (29)	46 (27)	59 (44)	58 (42)	59 (44)	58 (43)	60 (47)	56 (38)	59 (44)	49 (33)	48 (32)	99 (98)						
VH10	1G07AVH10	54 (40)	56 (41)	61 (54)	63 (57)	48 (36)	47 (36)	60 (55)	60 (48)	58 (48)	57 (48)	60 (52)	52 (40)	53 (42)	50 (38)	50 (38)	63 (53)	62 (52)					
	2G01AVH10	54 (38)	56 (42)	60 (54)	63 (56)	47 (35)	48 (34)	59 (57)	59 (49)	59 (49)	56 (45)	60 (51)	52 (39)	53 (42)	49 (39)	50 (39)	63 (54)	63 (53)	97 (94)				
VH11	1F10AVH11	58 (51)	61 (56)	54 (46)	56 (48)	56 (40)	56 (42)	54 (44)	52 (39)	50 (41)	52 (43)	52 (45)	53 (46)	56 (49)	56 (43)	55 (47)	50 (42)	50 (41)	51 (42)	51 (43)			
	2B03AVH11	58 (51)	61 (57)	53 (46)	55 (49)	55 (40)	55 (38)	53 (44)	51 (38)	51 (40)	51 (43)	51 (43)	52 (45)	57 (47)	55 (43)	56 (47)	49 (40)	49 (39)	51 (42)	51 (41)	97 (95)		
VH12	2G09AVH12	52 (38)	53 (44)	55 (45)	58 (49)	47 (35)	45 (30)	59 (52)	58 (50)	56 (49)	57 (46)	59 (47)	53 (41)	54 (44)	45 (35)	46 (34)	65 (51)	65 (50)	64 (57)	64 (56)	47 (38)	48 (38)	
VH13	2G11AVH13	59 (45)	61 (47)	48 (40)	48 (40)	62 (43)	60 (42)	49 (36)	48 (38)	48 (39)	49 (41)	51 (41)	58 (44)	55 (44)	57 (44)	60 (44)	50 (34)	50 (33)	50 (36)	49 (34)	54 (34)	55 (45)	51 (37)

Fig. 3. The nucleotide and amino acid similarities (shown in parentheses) between the VH-coding regions defined in representative clones from 13 different channel catfish VH families. The sequences were compared from the beginning of the FR1 region through the end of the FR3 region as aligned in Figs. 1 and 2. The number of sequence identities was used to calculate the pairwise percent similarity (sequence identities × 100/maximum number of residues compared).

Table 3
Comparison of nucleotide and amino acid similarity in the FR and CDR regions between sequences representing the 13 different channel catfish VH families

Region compared	Number of codons	Similarity range	Average similarity
FR1	28–29	38.0–67.8 (20.7–58.6)	51.3 ± 6.3 (39.3 ± 9.4)
FR2	14	50.0–92.9 (42.9–92.9)	70.2 ± 8.4 (66.8 ± 10.8)
FR3	32	45.6–86.7 (30.0–76.7)	61.1 ± 9.3 (52.8 ± 12.7)
FR total	74–75	48.4–74.9 (34.2–68.5)	59.0 ± 6.0 (50.1 ± 7.8)
CDR1	3–9	0.0–73.3 (0.0–80.0)	34.6 ± 19.0 (20.0 ± 21.1)
CDR2	15–18	12.5–66.7 (0.0–56.3)	41.7 ± 10.0 (22.5 ± 13.2)
CDR total	19–25	25.0–60.9 (0.0–48.0)	39.3 ±7.8 (21.1 ± 10.4)
Total	93–100	45.5–70.7 (29.9–64.0)	55.2 ± 5.6 (44.1 ± 7.3)

The percentage of nucleotide and predicted amino acid (in parentheses) similarity were determined from the aligned sequences shown in Figs. 1 and 2. The prototype sequences used for comparisons of the VH1–VH13 catfish families were: NG70, NG41, NG21, NG10, NG66, 10.4, 1C09AVH7, 3G01AVH8, 1G11AVH9, 1G07AVH10, 1F10AVH11, 2G09AVH12 and 2G11AVH13, respectively. The number of sequence identities was used to calculate the pairwise percent similarity (sequence identities × 100/maximum number of residues compared). The range and average percent similarity with the standard deviation for each specified region was calculated from the 78 pairwise comparisons that were determined. The range and average similarities for the three combined FR (FR total), the two combined CDRs (CDR total), and the total coding region similarities (not including the leader sequences) are indicated.

3.3. Germline VH segments representing members of the VH8–VH13 families

Critical to the definition and characterization of new VH gene families is the definition of genomic members of these families. Without this evidence there would be no reason to assume that these new groups of VH sequences did not arise by gene conversion or alternative recombination mechanisms involving VH members from known families. Accordingly, a channel catfish genomic library constructed in lambda DASH II was analyzed by different approaches, and positive clones which contained a member of the VH8–VH13 families were obtained. Following cloning of the requisite restriction fragments, the sequences of members representing these new VH gene families were defined and designated as VH8.1–VH13.1, respectively. The sequence was determined to minimally include 200 bp upstream from the leader initiation codon and was extended downstream through the RSS. The sequence of these germline VH segments is shown in Fig. 4 and is aligned with the prototype cDNA sequence assigned to the same family. The total nucleotide FR1–FR3 similarity between the aligned germline and cDNA sequences shown in Fig. 3 was greater than 90% (the alignment of VH11.1 terminated within the FR3 region). Therefore, these analyses confirmed that VH8–VH13 represent new VH families in the channel catfish.

The germline sequences VH8.1, VH9.1, and VH12.1 appear to be functional based upon the presence of upstream regulatory elements, a leader exon split by a short intron, an uninterrupted open reading frame corresponding to the V coding region, and a downstream RSS. In contrast, VH10.1, VH11.1, and VH13.1 appear to represent pseudogenes. In VH10.1 the leader exon has apparently been deleted; the sequence upstream of the coding region exon was extended about 300 bp and no leader sequence similar to that expressed in cDNA was identified. The VH10.1 coding region exon, however, encoded a single open reading frame that differed from 1G07AVH10 in 10 nucleotide positions (six predicted amino acid differences). In VH11.1 there have been multiple deletion events (a single bp deletion within the leader exon, a 5 bp deletion within the FR1-encoding region, and a major deletion within the FR3-encoding region that truncated the remaining downstream coding region and RSS) which rendered this gene non-functional. In VH13.1 the leader intron has a nonconsensus RNA acceptor splice site (GG versus AG) and a 2 bp deletion within the FR3 encoding region. This deletion results in a frameshift which changes the conserved terminal Y–Y–C–A–R into L–L–C–T–K. This change, which likely results in a nonfunctional structure, nonetheless maintains the open reading frame which may allow this gene to be expressed.

Potential transcriptional regulatory elements were located upstream from the leader sequences of each of the germline genes. These elements included an octamer and a TATA box. In each of these genes except for VH12.1, the octamer was in the mammalian H chain orientation and was either identical to the consensus (5′-ATGCAAAT) or differed by one nucleotide. The VH12.1 and VH13.1 sequences both had an additional upstream octamer motif which differed by two nucleotides from consensus. A potential TATA box is located 9–49 bp downstream of the octamer in each of these VH genes. The alignment with the 5’-ends of cDNA begin 25–28 bp downstream of the TATA box except in VH12.1 where this distance is 17 bp (Fig. 4).

The structure of the RSS within these catfish VH genes were generally conserved with a heptamer (CACAGTG) and a nonamer (ACAAAAACA) separated by a 22–24 bp spacer region (Fig. 5). The heptamer of the RSS exhibited the greatest variation in the fifth position (C rather than G), while pyrimidine substitutions in the ninth position of the nonamer were evident. These alignments also indicate that the spacer region has generally conserved motifs located at its 5’- and 3’-ends. The presence of phylogenetically conserved motifs

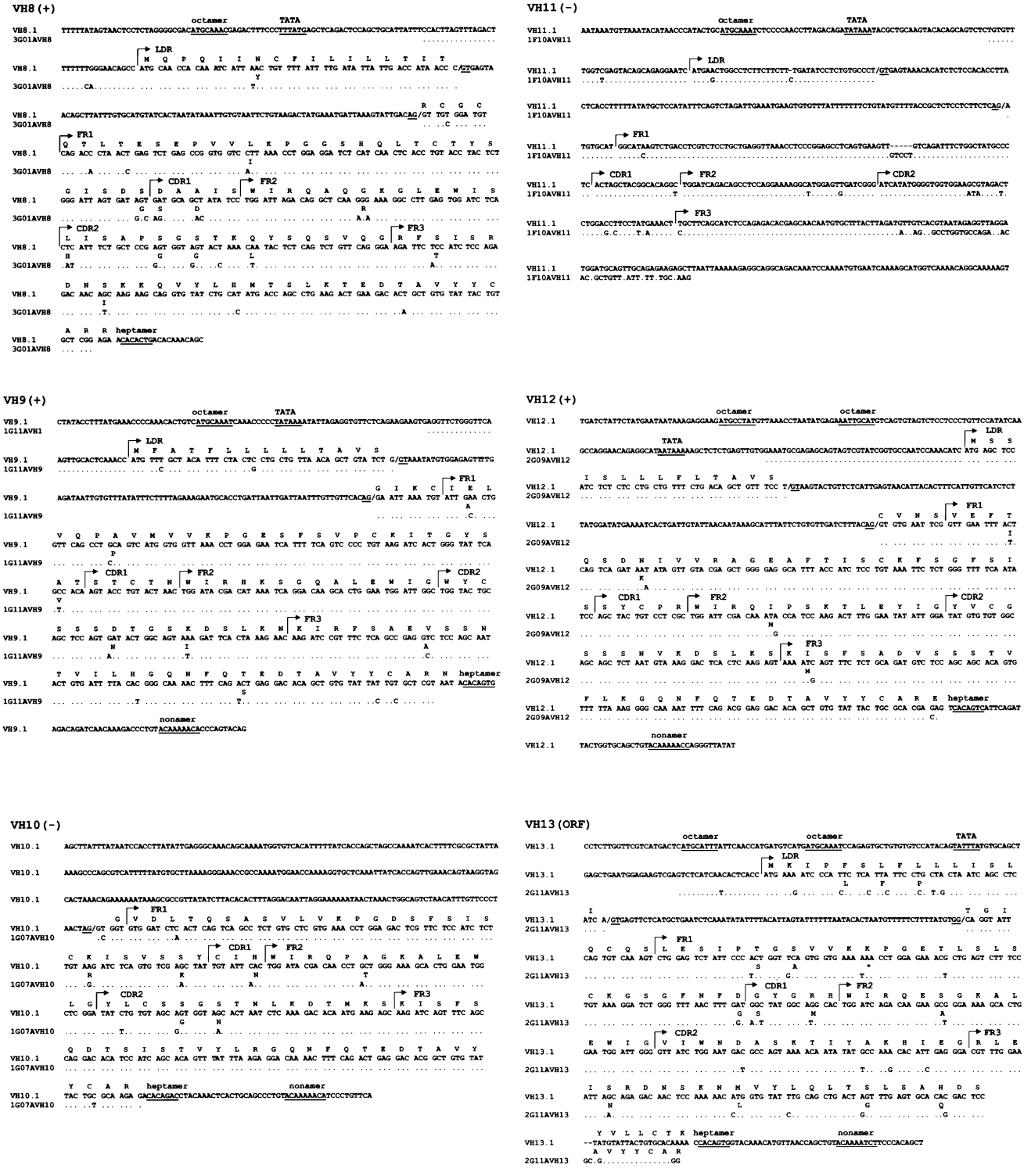

Fig. 4. Sequences of germline VH gene segments representing families VH8–VH13. The germline sequences are shown aligned with the prototype cDNA sequences representing the same VH family. A (+) next to the VH family designation indicates that the gene is likely functional; a (−) indicates that the gene is a pseudogene. The ORF designation for VH13 indicates that this gene contains a single open reading frame but in comparison to cDNA this gene has suffered a two bp deletion within the encoding FR3 which likely renders this gene nonfunctional. Putative regulatory octamer and TATA box sequences, the splice donor and splice acceptor sites within the leader intron, and the heptamer and nonamer of the recombination signal sequence are indicated and underlined. Nucleotide identity is indicated by dots whereas gaps are indicated by dashes. The predicted amino acid sequence encoded by the functional VH segments is shown.

	FR3						heptamer		nonamer	
CatfishVH1.1	TAT	TAC	TGT	GCA	CGA	CACT-	CACACTG	ACACAACAGA--CTCCGAGACTGT	ACAAAAACA	TACACAAGCACGTCTCAGAGC
CatfishVH3.1	...	...	..C	..C	..T	G..C-	....G..	.T..TT.TCC--..AGACAT.A..	.....	C.TG.TT..TTTAG...C.TG
CatfishVH4.1	...	...	..C	A.C	..C	...A-	.....	.TG.TT.TAC--T.GG.T.T.AA.	.....	.GTG.GT.AT.C..A..C.CT
CatfishVH6.1	...	...	..C	..T	..T	G.AC-	....G..	.G...GATC.ACAA.ATCCC....	.T..GG...	CC..GT.CAGAAAACTCTTAA
CatfishVH7.1	...	...	C.C	..C	...	G.AGC	....G..	..TGG.ACTGA-GGAAACAG..T.	..TCG...T	ACGCAC.TATTT.TCACACAG
CatfishVH8.1	...	...	...	..T	..G	AGAA-	.....	.....ACAGC----------	---------	----------
CatfishVH9.1	...	..T	...	..T	..T	A.TA-	....G..	.G...GATC.ACAAAGA-CC....	.....	CC..GT.CAGAAGA.T-----
CatfishVH10.1	...	...	..C	...	A..	G.---	....GAC	CT....ACTCA-..G.AGCC....	.....	.C.CTGTT..TT.TC.TTTCT
CatfishVH12.1	...	...	..C	...	...	G.G.-	....G.C	.TT..GATT.C-TGGT.CAG....	C	AGGGTT.TATTA.G.TGAGTT
CatfishVH13.1	...	...	...	...	.A.	A..--	....G..	GT....ACATG-T.AACCAG....	.....T.T	.C.CAC...T.TCA.A.A..A
FuguVHI	...	..T	...	..T	...	G.G.-	....G..	..TG..GTC.G-T..A.CAG....	.....	C....TTCTCATAT.T.CT..
FuguVHII	...	...	...	..C	G..	..---	....A.A	.....ACC.T-.AGT.GAG...A	C	CCTCAG....TCAACACA.CA
TroutVH1	...	..T	...	..C	A..	G.---	....G..	.....GG..G--..G.AGCG....	.....T	----------
GoldfishVH1	...	...	...	..G	A..	G.---	....G..	.AGAG.A.CTG-.GAGTG.C.AA.	C	C.AC.TGCAG----------
GoldfishVH2	...	...	...	...	A..	G...-	...TG.T	.....A...--..G.AG...A	.....T	ATTCATGCAG.AG--------
SalmonVH5	...	...	...	..C	..G	T..A-	....G.A	.G.G.GAGTGG-TGGGAGAG....	.T.....C	ACTT.CCCATG.AGCACAGA.
DanioVH1	...	...	...	...	..G	ACT--	...GG.T	G.....A.C.--..G.AG.....A	.....	ATGCATGCAGAAAA...CT.A
HumanVH2	...	...	...	...	.AC	AGAC-	....AA.	.....G.CC.-GGG.ACCTC....	C	C.GG.TGCTT.TCA.TG.T..
Consensus	TAT	TAC	TGT	GCV	CGA	GANN-	CACAGTG	ACACAAAHNAN-YTRNRSVNCTGT	ACAAAAACA	

Fig. 5. Alignment of the recombination signal sequence (RSS) and adjacent flanking regions from VH segments representing different VH families from different species of bony fish. The catfish VH sequences from families VH1–VH7 are from previous studies (Ventura-Holman et al., 1994, 1996) whereas the VH sequences from families VH8–VH13 are from the present study (members from families VH2, VH5, and VH11 are not shown because the RSS was either deleted or truncated in presently sequenced members of these families). Other germline teleost VH sequences and their accession numbers were obtained from the IMGT database (http://www.imgt.cines.fr) and represented the following species: zebrafish (*Danio rerio*, VH1, AF273897), fugu (*Takifugu rubripes*, VHI and VHII, both from AF108421), rainbow trout (*Oncorhyncus mykiss*, VH1, M57442), goldfish (*Carassius auratus*, VH1, J03616; VH2, X61312), and the Atlantic salmon (*Salmo salar*, VH5, Y12452). A representative germline member of the human VH2 family (VH2-5, AB019440) is shown because the consensus RSS spacer sequence within members of this family share similarity to that found in these various species of bony fish. The demarcation of the FR3-encoding region, the heptamer, spacer, nonamer, and 3′-flanking region is shown. Nucleotide identities are indicated by dots, whereas gaps introduced to maximize similarities are indicated by dashes. The conserved nucleotides with the RSS spacer region are underlined.

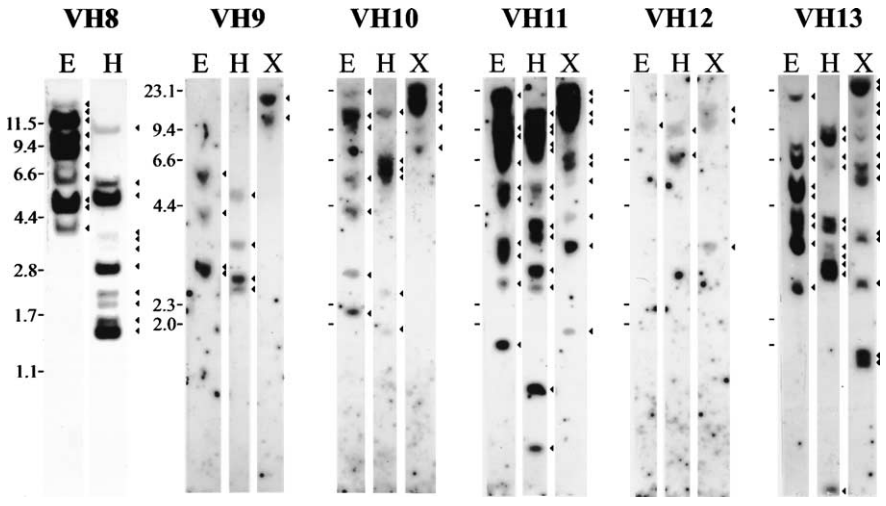

Fig. 6. Southern blot analysis of restricted genomic DNA hybridized with VH8–VH13 family specific probes. The DNA was restricted with the following enzymes: E, *Eco*RI; H, *Hind*III; X, *Xba*I. The sizes of the hybridizing bands shown in kilobases were measured by using *Pst*I- or *Hin*dIII-digested lambda DNA. The position of each hybridizing band is indicated by a blackened triangle.

within the RSS spacer region suggests that they may be important in gene recombination (see Section 4).

Genomic Southern blot analyses were used to determine the number of closely-related germline genes in each of these new VH families. Genomic DNA was isolated from the nucleated erythrocytes of individual catfish, restricted with *Eco*RI, *Hin*dIII, or *Xba*I, and blots were hybridized under high stringency conditions with probes specific for each VH family. These results identified the following number of different sized restriction fragments: VH8, 10–11; VH9, 4; VH10, 6–7; VH11, 10–13; VH12, 2–3; and VH13, 9–14 (Fig. 6). If it is assumed that the hybridizing fragments contain only one member, and that co-migration of similarly sized fragments does not occur, then these new families range from small to medium in size when compared to the estimated number of members in the other VH families (VH1, 22–28; VH2, 20–24; VH3, 17–21; VH4, 11–15; VH5, 28–32; VH6, 17–20; and VH7, 8–10; Ghaffari and Lobb, 1991; Warr et al., 1991; Ventura-Holman et al., 1994). Therefore, these combined analyses estimate that there are about 165–200 germline VH gene segments represented by 13 different VH families in the channel catfish.

4. Discussion

These studies have defined six new VH gene families in the channel catfish, which when coupled with earlier analyses, allows certain conclusions to be made regarding the structure, genomic organization, and phylogeny of VH diversity in early vertebrate evolution. The channel catfish has 13 VH gene families, designated VH1–VH13. This is the highest number of VH families presently defined in a lower vertebrate. Sequence comparisons indicate there is extensive diversity between members of these different families; the average nucleotide similarity between representative members of these different families is 56% (range 32–64%), with the greatest variability located within the CDR-encoding regions. The number of members within these individual families ranges from very few (two or three members as in VH12) to very large (20 or more members as in families VH1, VH2, and VH5). Genomic Southern blot analyses estimate that the number of genomic segments represented by these 13 families ranges between 165 and 200 if it is assumed that co-migration of similar sized restriction fragments does not occur, that each band contains only a single VH gene, and that restriction sites identical to those used in the genomic digests are not located within the VH segments. However, it is clear from this, as well as previous studies, that a significant proportion of catfish VH germline genes are pseudogenes. At this point about 10% of the germline VH genes in catfish have been sequenced, and these analyses indicate that approximately 50% are pseudogenes due to either insertion/deletion events within the leader or the coding region, or mutations within the RSS. If this percentage is applied to the estimated number of germline VH segments, then there may be as many as 80–100 functional genomic VH segments represented by these 13 VH families. These combined results indicate that VH genomic diversity in the catfish is extensive; this level of diversity appears to be greater than that found in man where only 44 of the 123 germline VH segments are functional, and these segments are represented by only seven different families (Matsuda et al., 1998).

Earlier studies on the genomic organization of the catfish VH segments had shown that members of the different VH gene families are closely-linked and extensively interspersed within the IgH locus. Present analyses support this conclusion. The VH8.1 segment was identified by sequencing a region adjacent to a VH7 germline segment. The genomic lambda clones that contained the germline genes VH10.1, VH11.1 and VH13.1 were shown by hybridization analyses to contain members representing at least five different VH gene families (Ventura-Holman et al., 1994). The genomic clones which contained VH9.1 and VH12.1 segments each contained members of one additional VH family. Based upon these results with these genomic clones, which contain an average insert size of 16–18 kb, VH segments are closely linked with an expected average distance between segments of approximately 3 kb.

The germline gene segments representing these new VH families shared common characteristic structural features: an upstream region that contained potential transcriptional regulatory elements, a leader sequence interrupted by a short intron, and the V coding region terminating with a RSS. The octamer is one of the most important regulatory elements of Ig gene transcription, and this motif is conserved in Ig promoters and enhancers of H and L chain genes (reviewed by Henderson and Calame, 1998). In the sequenced gene segments representing these 13 VH families, an octamer in the mammalian H chain orientation (5′-ATGCAAAT) was identified upstream of the likely TATA box. The octamer motifs were either identical to the consensus motif or had one or two mismatches. The distances between the octamer and the leader initiation codon ranged between 61 and 110 nucleotides in the functional germline segments. The identified members in families VH4, VH12, and VH13 each had additional octamer motifs upstream of the likely TATA box. VH12.1 was the only one of these sequences to have this additional octamer with one mismatch from consensus in the reverse or L chain orientation. Database searches of transcription factor binding consensus sites revealed no other motifs in the region located between the octamer and leader initiation codon that were generally conserved in these members of the different VH families.

The leader sequence, which encodes 13–17 amino acids, is separated by a short intron which varies in length between 82 and 119 bp in the functional VH segments from these families. There is no intron sequence similarity between members of these different families. The coding region is readily delineated into the encoding FR and CDR regions based upon sequence comparisons with higher vertebrate VH genes (Kabat et al., 1991). However, the presence of

cysteine within CDR regions within members of VH6, VH9, VH10, and VH12 is unusual. Cysteine was encoded within both the CDR1 and CDR2 in expressed members of these families, and perhaps even more important was that cysteine was encoded in the CDR1 and CDR2 regions of germline VH segments VH6.1, VH9.1 VH10.1, and VH12.1. Therefore, the presence of cysteine residues within CDR regions can not be simply attributed to potential somatic mutation events which occurred within these expressed sequences. The presence of cysteine within CDR regions, although not typical of human or mouse V regions, has been observed in V regions of other species. In cattle the CDR3 typically contains six to eight cysteine residues which have been suggested to form intra-CDR3 disulfide bridges which may stabilize the extremely long CDR3 (Saini et al., 1999; Ramsland et al., 2001). H chain cDNA from the platypus also encodes additional cysteine residues which are located within the CDR2 and the CDR3 or are located just within the CDR3 (Johansson et al., 2002). Shark NAR receptors also contain noncanonical cysteine residues within FR regions or are encoded by expressed D segments within CDR3 regions (Roux et al., 1998). In camelids the V_HH domain acquires a bonafide disulfide bond between cysteine residues in CDR1 and CDR3 or between cysteines in CDR3 and FR position 45. The structures of camelid V_HH antibodies in complex with antigen indicates that the structures adopt the classical Ig fold; however, the additional disulfide bridges provide noncanonical CDR conformations which allow for dispersal or protrusion of CDR regions as well as significant interactions of FR residues with bound antigen (Desmyter et al., 2002). These studies suggest that cysteine residues in catfish CDR1 and CDR2 regions may also form a disulfide bridge that may alter CDR conformations, and equally importantly suggest that FR regions, which are highly diversified in these four different VH families, may play an important role in antibody-antigen interactions. Studies are underway to determine sequence variability within the CDR regions in members of these families which should provide the support necessary to make further conclusions.

In earlier sequence comparisons of the RSS in human VH genes, Hassanin et al., 2000, had found that the RSS could be grouped into four types based upon similarities within the spacer and 3′ flanking regions. We compared the RSS and 3′ flanking regions in members of the different catfish VH families as well as other available germline sequences from other bony fish with the human RSS groups. These analyses indicated that there was no similarity within the RSS 3′-flanking region; however, there were conserved motifs located at the 5′- and 3′-ends of the spacer region. The 5′-motif is ACACAAA, and the 3′-motif is CTGT. These motifs share striking resemblance to motifs internal to the consensus sequences for the nonamer and heptamer. Studies have indicated that the sequence of the spacer region may influence the frequency of gene recombination, although the spacer does not appear to affect cleavage rates of recombination substrates by the recombinase enzymes (Nadel et al., 1998; Yu et al., 2002). Recent RSS footprinting studies have also shown that different regions within the spacer sequence are protected when pre- and post-cleavage synaptic complexes are compared (Nagawa et al., 2002). These analyses indicate that nucleotides within the spacer region may influence the binding of the recombinases and hence possibly account for the sequence conservation in the 5′- and 3′-ends of the RSS spacer region in these phylogenetically diverse species.

Phylogenetic studies by Ota and Nei (1994) had classified VH genes of different vertebrates into five major groups designated A–E, with Groups A–C containing clans I–III of human VH genes (Kirkham et al., 1992). Their analyses with 55 sequences indicated that Groups A and B each contained VH genes from both mammals and *Xenopus*; Group C included VH genes from mammals, toad, chicken, caiman, coelacanth, and bony fish; Group D represented VH genes found in only bony fish; and Group E represented VH genes found only in the cartilaginous fish. With the characterization of 13 VH families in the catfish and cDNA studies in rainbow trout identifying 11 VH families (Roman et al., 1996), we re-examined the phylogenetic relationships of VH genes to determine if additional insight could be defined. We patterned these analyses after the methods used by Ota and Nei (1994) and conducted phylogenetic analyses with 59 VH sequences representing members of different VH families using the minimum evolution method with Poisson correction (Rzhetsky and Nei, 1992). The reliability of the tree obtained was tested by either bootstrap resampling or the interior branch length test, with the results from both approaches supporting the major branches of the tree topology shown in Fig. 7. The major branches of the tree show the same five major groups of VH genes, A–E. It is, however, important to note that the branch length confidence values do not support specific inferences into the evolutionary relationships of the main branches A–E. The limiting feature in these analyses is that these comparisons are based upon only 77 positions (the combined amino acid FR regions), and this limited number of positions allows for higher stochastic error than would be the case if more positions could be evaluated.

These analyses indicate that the catfish VH families and the trout VH families cluster into either Group C, the only group which contains sequences representing different VH families from species across major phylogenetic lineages, or Group D. Therefore, none of these families can be placed into Groups A, B, or E. These analyses also indicate a close phylogenetic relationship between some VH families in the catfish and the trout, with the highest confidence levels within the tree defined between catfish VH1 and trout VH9, between catfish VH7 and trout VH4, and between catfish VH11 and trout VH11. Pairwise nucleotide sequence comparisons between members in these pairs of families (extending from FR1 through the end of FR3 including CDR1 and CDR2) indicated that the overall total sequence similarity was less than 80% (range 63.6–79.7%). The alignment

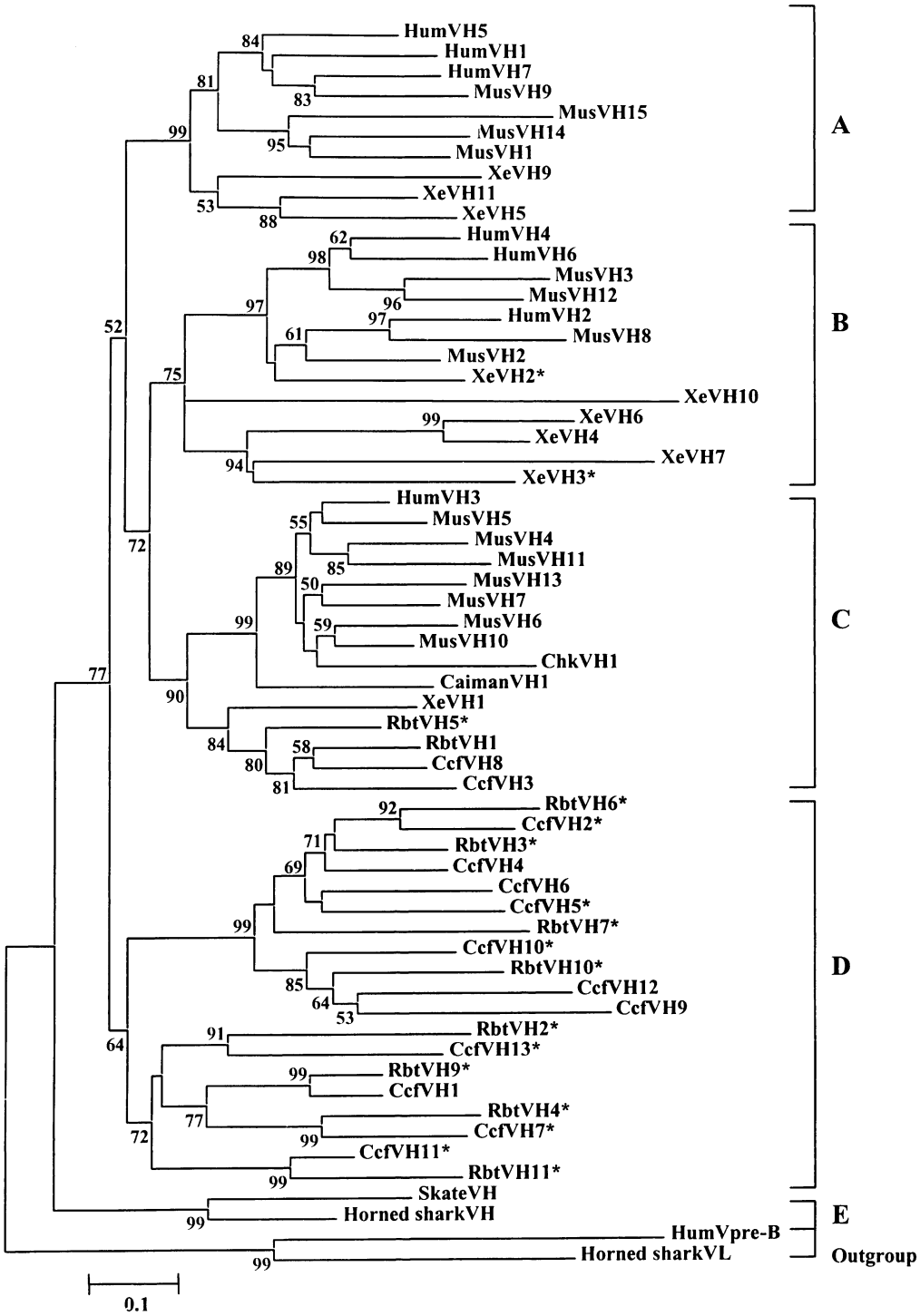

Fig. 7. Phylogenetic tree of VH sequences representing different VH families from divergent lineages of vertebrates. Fifty-nine sequences representing different VH families and two outgroup sequences were analyzed using the minimum evolution method with Poisson correction (Rzhetsky and Nei, 1992). Germline VH sequences were used if available, otherwise cDNA sequences (represented by an asterisk) were used. The numbers shown reflect the confidence probability of each interior branch (<50% not shown). The interior branch lengths are measured in terms of the number of amino acid substitutions with the scale given below the tree. The names and sequences used in these analyses are indicated in Table 2 of the text. The five major groups of VH genes are designated with the letters A–E.

of NG70 (VH1) with trout X81507 (VH9) yielded the highest overall nucleotide similarity of 79.7%. This value is very close to the numerical definition of members within the same family; as additional members in these two VH families are defined, the relative overlap between these closely-related gene families will become clearer.

Secondly, Group D in all tree topologies that we examined is represented by two major groups. One of these groups is statistically closely related, while the other, although less significant, still forms a separate branch within Group D. This relationship was not clearly established until these additional VH families were identified. These two branches within Group D contain families from both the catfish and the trout, and additional comparisons using the presently limited information available on different VH families in other species of bony fish supported this conclusion (additional trees not shown). The major lineages of teleost bony fish likely diverged from a common ancestor during the Cretaceous period more than 100–150 million years ago (Colbert, 1969). Therefore, these analyses indicate that three major lineages of VH families arose early in the phylogeny of the bony fish: one lineage (Group C) was sustained during vertebrate phylogeny, while two lineages (both within Group D) appear to be present only within the bony fish. These results suggest to us that structural divergence likely evolved to provide major functional differences between these VH lineages.

In conclusion, these studies show that VH genes diverged early in phylogeny into complex multiple gene families that retained basic characteristic structural features. These features include the basic two exon germline structure with transcriptional regulatory elements located immediately upstream from the leader initiation codon within the first exon and the open reading frame of the second exon terminating in the phylogenetically conserved RSS necessary for gene recombination. Within this format VH genes diverged FR sequences that are distinct from one VH family to another while preserving amino acid residues critical for proper Ig domain folding. Accompanying this divergence were major changes within the CDR regions, most of which appear to be family specific, which affected the composition and length of these regions critical in antigen-antibody interactions.

Acknowledgements

This work was supported by grant AI23052 from the National Institutes of Health. The authors thank Mary Duke for high throughput cDNA sequencing support.

References

Bengten, E., Quiniou, M.-A., Stuge, T.B., Katagiri, T., Miller, N.W., Clem, L.W., Warr, G.W., Wilson, M., 2002. The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. J. Immunol. 169, 2488–2497.

Brodeur, P.H., Osman, G.E., Mackle, J.J., Lalor, T.M., 1988. The organization of the mouse IgH-V locus: dispersion, interspersion, and the evolution of V_H gene family clusters. J. Exp. Med. 168, 2261–2278.

Colbert, E.H., 1969. Evolution of the Vertebrates: A History of the Backboned Animals Through Time. Wiley, New York, pp. 10–63.

Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., Cambillau, C., 2002. Three camelid V_HH domains in complex with porcine pancreatic α-amylase. Inhibition and versatility of binding topology. J. Biol. Chem. 277, 23645–23650.

Ghaffari, S.H., Lobb, C.J., 1989a. Cloning and sequence analysis of channel catfish heavy chain cDNA indicate phylogenetic diversity within the IgM immunoglobulin family. J. Immunol. 142, 1356–1365.

Ghaffari, S.H., Lobb, C.J., 1989b. Nucleotide sequence of channel catfish heavy chain cDNA and genomic blot analyses. Implications for the phylogeny of Ig heavy chains. J. Immunol. 143, 2730–2739.

Ghaffari, S.H., Lobb, C.J., 1991. Heavy chain variable region gene families evolved early in phylogeny: Ig complexity in fish. J. Immunol. 146, 1037–1046.

Ghaffari, S.H., Lobb, C.J., 1992. Organization of immunoglobulin heavy chain constant and joining region genes in the channel catfish. Mol. Immunol. 29, 151–159.

Ghaffari, S.H., Lobb, C.J., 1999. Structure and genomic organization of a second cluster of immunoglobulin heavy chain gene segments in the channel catfish. J. Immunol. 162, 1519–1529.

Hassanin, A., Golub, R., Lewis, S.M., Wu, G.E., 2000. Evolution of the recombination signal sequences in the Ig heavy-chain variable region locus of mammals. Proc. Natl. Acad. Sci. U.S.A. 97, 11415–11420.

Hayman, J.R., Lobb, C.J., 2000. Heavy chain diversity region segments of the channel catfish: structure, organization, expression, and phylogenetic implications. J. Immunol. 164, 1916–1924.

Hayman, J.R., Ghaffari, S.H., Lobb, C.J., 1993. Heavy chain joining region segments of the channel catfish. Genomic organization and phylogenetic implications. J. Immunol. 151, 3587–3596.

Henderson, A., Calame, K., 1998. Transcriptional regulation during B cell development. Annu. Rev. Immunol. 16, 163–200.

Johansson, J., Aveskogh, M., Munday, B., Hellman, L., 2002. Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. J. Immunol. 168, 5155–5162.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., Foeller, C., 1991. Sequences of Proteins of Immunological Interest. National Institutes of Health, Bethesda, MD.

Kirkham, P.M., Mortari, F., Newton, J.A., Schroeder Jr., H.W., 1992. Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. EMBO J. 11, 603–609.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Lobb, C.J., 1985. Covalent structure and affinity of channel catfish anti-dinitrophenyl antibodies. Mol. Immunol. 22, 993–999.

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H., Miyata, T., Honjo, T., 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188, 2151–2162.

Max, E.E., 1999. Immunoglobulins: molecular genetics. In: Paul, W.E. (Ed.), Fundamental Immunology. Lippincott–Raven, Philadelphia, pp. 111–182.

Nagawa, F., Kodama, M., Nishihara, T., Ishiguro, K., Sakano, H., 2002. Footprint analysis of recombination signal sequences in the 12/23 synaptic complex of V–D–J recombination. Mol. Cell. Biol. 22, 7217–7225.

Nadel, B., Tang, A., Escuro, G., Lugo, G., Feeney, A.J., 1998. Sequence of the spacer in the recombination signal sequence affects V–D–J rearrangement frequency and correlates with nonrandom *Vκ* usage in vivo. J. Exp. Med. 187, 1495–1503.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, pp. 87–113.

Ota, T., Nei, M., 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol. Biol. E 11, 469–482.

Ramsland, P.A., Kaushik, A., Marchalonis, J.J., Edmundson, A.B., 2001. Incorporation of long CDR3s into V domains: implication for the structural evolution of the antibody-combining site. Exp. Clin. Immunogenet. 18, 176–198.

Roman, T., Andersson, E., Bengten, E., Hansen, J., Kaattari, S., Pilstrom, L., Charlemagne, J., Matsunaga, T., 1996. Unified nomenclature of Ig VH genes in rainbow trout (*Oncorhynchus mykiss*): definition of eleven VH families. Immunogenetics 43, 325–326.

Roux, K.H., Greenberg, A.S., Greene, L., Strelets, L., Avila, D., McKinney, E.C., Flajnik, M.F., 1998. Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc. Natl. Acad. Sci. U.S.A. 95, 11804–11809.

Rzhetsky, A., Nei, M., 1992. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J. Mol. E 4, 367–375.

Saini, S.S., Allore, B., Jacobs, R.M., Kaushik, A., 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. Eur. J. Immunol. 29, 2420–2426.

Ventura-Holman, T., Lobb, C.J., 2002. Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters. Mol. Immunol. 38, 557–564.

Ventura-Holman, T., Jones, J.C., Ghaffari, S.H., Lobb, C.J., 1994. Structure and genomic organization of VH gene segments in the channel catfish: members of different VH gene families are interspersed and closely linked. Mol. Immunol. 31, 823–832.

Ventura-Holman, T., Ghaffari, S.H., Lobb, C.J., 1996. Characterization of a seventh family of immunoglobulin heavy chain VH gene segments in the channel catfish, *Ictalurus punctatus*. Eur. J. Immunogenet. 23, 7–14.

Warr, G.W., Middleton, D.L., Miller, N.W., Clem, L.W., Wilson, M.R., 1991. An additional family of VH sequences in the channel catfish. Eur. J. Immunogenet. 18, 393–397.

Yu, K., Taghva, A., Lieber, M.R., 2002. The cleavage efficiency of the human immunoglobulin heavy chain VH elements by the RAG complex. J. Biol. Chem. 277, 5040–5046.